SCIENTIFIC DISCOVERY AS PROBLEM SOLVING

Technical Report AIP - 101

Herbert A. Simon

Carnegie Mellon University
Department of Psychology

19 Feb 1989

The Artificial Intelligence and Psychology Project

Departments of
Computer Science and Psychology
Carnegie Mellon University

Learning Research and Development Center
University of Pittsburgh

2

SCIENTIFIC DISCOVERY AS PROBLEM SOLVING

Technical Report AIP - 101

Herbert A. Simon

Carnegie Mellon University
Department of Psychology

19 Feb 1989

This research was supported by the Defense Advanced Research Projects Agency, Department of Defense, ARPA Order 3597, monitored by the Air Force Avionics Laboratory under contract F33615-81-K-1539, and by the Computer Sciences Division, Office of Naval Research, under contract number N00014-86-K-0678.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AIP - 101	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Carnegie-Mellon University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Computer Sciences Division Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) Department of Psychology Pittsburgh, Pennsylvania 15213	7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Same as Monitoring Organization 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0678
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS p4000ub201/7-4-86 PROGRAM ELEMENT NO: N/A; PROJECT NO: N/A; TASK NO.: N/A; WORK UNIT ACCESSION NO: N/A

11. TITLE (Include Security Classification)

Scientific Discovery as Problem Solving

12. PERSONAL AUTHOR(S)

Herbert A. Simon

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	89/2/19	22

16. SUPPLEMENTARY NOTATION

Submitted to RISESST Journal, Torino, Italy, February 1989

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	scientific discovery, problem solving, physical symbol system, computer simulation, explanatory theories, insight

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

ABSTRACT

The quick and highly incomplete account of the evidence now available about discovery processes confirms both the variety and heterogeneity of the activities that make up the enterprise of science and the consistency with which these activities conform to the pattern of heuristic search. Such phenomena as intuition, insight, and inspiration derive from the capacity for recognition that every expert acquires in the domain of expertise.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Alan L. Meyrowitz	22b. TELEPHONE (Include Area Code) (202) 696-4302 22c. OFFICE SYMBOL N00014

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Scientific Discovery as Problem Solving

Herbert A. Simon
Carnegie-Mellon University

Scientific discovery is a central topic both for the philosophy of science and for the history of science, but it has been treated in very different ways by the two disciplines.[1] Philosophers of science have been concerned mainly with the verification (or falsification) of scientific theories rather than with the origins of the theories or the processes by means of which they were derived. In fact, until very recently, most philosophers of science denied that a theory of discovery processes was possible, and that is probably still the majority view in the discipline (Popper, 1961). Within the past decade or so, however, a weak trickle of interest in discovery has grown into a sizeable stream (Nickles, 1978).

Historians of science, on the other hand, have long been interested in the discovery process, and have experienced no special difficulty in studying it. I do not mean by this that it is easy to study. Since data are mostly lacking at the more microscopic level -- the hour by hour progress of an investigation -- they have viewed it on a more global scale, usually relying on publications as a principal source of data. Sometimes, however, and especially in recent years, they have gained access to more detailed accounts of scientific work: for example the diaries and correspondence of Darwin (Gruber, 1974), and the laboratory notebooks of Faraday (1932-36) and Hans Krebs (Holmes, 1980). These sources allow a discovery to be traced, if not minute by minute, at least experiment by experiment.

Cognitive psychology is a third discipline that has long had an interest in

---

[1] In this paper, I have drawn extensively upon *Scientific Discovery: Computational Explorations of the Creative Process*, Cambridge, MA: MIT Press, 1987, an account of research carried out over a decade with my co-authors Pat Langley, Gary L. Bradshaw, and Jan M. Zytkow. Of course, I alone am responsible for this particular expression of our views, but the ideas are a wholly joint product.

scientific discovery. Since psychology is especially concerned with thinking at the micro-level, progress toward understanding the discovery process has been impeded by the absence of appropriate data. To make up for this lack, psychologists have sometimes had to rely upon anecdotal accounts of discovery, never wholly reliable, or upon experiments which elicited behaviors that might or might not be correctly regarded as creative (Glass, Holyoak, & Santa, 1979, pp. 432-440).

The development of modern cognitive science, combining research methods drawn from psychology and artificial intelligence, has produced a new burst of activity in the study of scientific discovery (Giere, 1988). The substantive innovation sparking this activity is a theory of human problem solving that has been constructed over the past thirty years which might be capable of accounting for scientific discovery as well as for more mundane kinds of human mental activity. The methodological innovation sparking the new research has been the use of electronic computers to simulate human thinking, and accordingly, the use of computer programs as theories (systems of difference equations) of thinking (Newell & Simon, 1972).

I will begin my account of these developments by trying to define what we would mean by a "theory of scientific discovery," what the shape or form of such a theory, descriptive or normative, might be. Next, I will describe the theory of human problem solving that has emerged from the research in modern cognitive psychology (or, to include the simulation aspects, cognitive science). With these preliminaries out of the way, I will say something about the theory of discovery that has emerged, its relation to the theory of problem solving and the evidence that supports it.

Shape of a Theory of Discovery

Arguments agains the possibility of constructing a theory of scientific discovery fall into two major categories. Stated simply, the first argument is that a successful

theory of discovery processes would predict discoveries, hence make them -- an obvious impossibility. The second argument is that discovering a theory requires a "creative" step, and that creativity is inherently unexplainable in terms of natural processes.

I will not spend much time in debating these points, for they can be refuted by a constructive proof. After one has described a giraffe to a skeptical listener and been assured that "there ain't no such animal" the best reply is to exhibit a living and breathing example. Computer programs exist today that, given the same initial conditions that confronted certain human scientists, remake the discoveries the scientists made. BACON, described at length in Langley, et al., is one such program; KEKADA (Kulkarni and Simon, 1988) is another. Neither BACON nor KEKADA has made (predicted?) a wholly new discovery, but there is no reason in principle why they should not. Other computer programs -- e.g., DENDRAL -- in fact have done so (Feigenbaum et al., 1971).

What do we mean by claiming that these programs are theories of scientific discovery? We mean, first, that a computer program is formally describable as a system of difference equations; hence has essentially the same logical structure as those theories in the physical and biological sciences that take the form of systems of differential or difference equations. For any given state of the system under study, such a theory predicts the subsequent state. We mean, second, that the symbolic processes that enable the programs to make discoveries, or to reproduce historical ones, can be shown empirically to resemble the processes used by human scientists. I will have more to say about this evidence presently.

Exhibiting "living and breathing" computer programs that actually make discoveries also refutes the argument that creativity cannot be explained or simulated. To maintain this claim in the face of the programs' accomplishments would imply that

Kepler, and Joseph Black, and Snell, and Dalton, and Avogadro, and Faraday, and Hans Krebs, and many others whose discovery processes have been simulated, were not creative. If they were not, then creativity is obviously not required for making scientific discoveries of first magnitude. If they were, then the creative processes have been simulated. Preferring the latter alternative, I will have more to say presently about the nature of creativity.

What kind of evidence would allow us to conclude that the processes in a computer program were basically the same as the processes used by a human being in the course of thinking? We have every reason to believe that while some thought processes are conscious, so that their inputs and outputs can be reliably reported by the thinker (Ericsson and Simon, 1984), other essential processes are subconscious and not open to direct observation or even self-observation. Since the unobservable processes will be as essential to the theory as the observable ones, we are faced with the necessity of inferring their presence by indirect means.

This difficulty is no different from those faced continually in all of the sciences. If, for example, the information processes that allow a person to recognize a familiar object are inaccessible to direct observation, as they are, it is equally true that electrons, atoms, and certainly quarks, are inaccessible to direct observation. The task of inducing theories from data and the task of persuading ourselves of the validity of these theories, is no different in the case before us than in any other domain of science.

The task of a theory of discovery is to postulate an organization of processes sufficient to account for the discoveries and for the observable behavioral phenomena that accompany them. The task of a theory of verification (I will take a Popperian position here, but a Bayesian one would do as well) is to test that the theory is not falsified by any of the observations of the relevant phenomena. In this formulation, a

theory of scientific discovery, like any other theory, is falsifiable but not irrevocably verifiable. The theory I shall propose here is a theory of both conscious and unconscious thought processes described at the level of information processes or symbolic processes. If it is false, that can be shown by demonstrating serious mismatches of the behavior of the program that embodies it with the behavior of human beings engaged in creative activity.

The Theory of Human Thinking

One claim implicit in BACON, KEKADA, and similar models of the discovery process is that the thought processes of scientists are basically the same as the human thought processes that have been modeled and simulated in more "mundane" task environments. The scientist does not think in ways that are qualitatively different from the ways in which other professionals think, or the ways in which college students think when confronted with puzzles to solve in the psychological laboratory, or the ways in which T. C. Mits (The Common Man In The Street) thinks.

Today we have empirically validated theories of the thought processes of chessplayers, of medical diagnosticians, of subjects solving the Tower of Hanoi and Missionaries and Cannibals puzzles, of students solving algebra and physics problems, and of thinkers and problem solvers in many other domains.[2] These theories, like the theories of scientific discovery that are under discussion here, take the form of computer programs that actually perform the tasks in question. Hence in each case they postulate a set of processes that is at least *sufficient* to perform the task. In addition, a large body of evidence, including evidence from thinking-aloud protocols and eye-movement records, shows that the processes are closely similar to those

[2]For reviews of some of the evidence, see Newell & Simon, 1972; or Simon, 1979, 1989).

used by human subjects.

A reassuring feature of these models is that they are closely similar in structure from task to task. What distinguishes thinking in one task domain from that in another is specific substantive knowledge of the domain, and not the processes used to apply this knowledge to the task. This commonality of process throughout the range of tasks that has been studied gives us initial confidence -- a high Bayesian prior -- that the same processes will also show up in the study of scientific discovery. What are the underlying common features of this theory of thinking?

Physical Symbol Systems

The foundation for all of the models is the Physical Symbol System Hypothesis (Newell & Simon, 1976). The hypothesis states that physical symbol systems, and only such systems, are capable of thinking. A *physical symbol* is a pattern (of chalk, ink, neuronal connections, electromagnetic fields, or what not) that refers to or designates another pattern or a detectable external stimulus. Printed words on a page are symbols, so are pictures or diagrams, so are numbers. A *physical symbol system* is a system that is capable of *inputting* symbols, *outputting* them, *storing* them in memory, *forming and modifying structures* of symbols in memory, *comparing* pairs of symbols for identity or difference, and *branching* in its subsequent behavior on the basis of the outcomes of such tests.

A computer is obviously a physical symbol system. Its ability to perform these processes (and only these processes) can be verified easily from its physical properties and operation. A human brain is (less obviously) a physical symbol system. It can certainly carry out the processes specified in the definition of such a system, but perhaps other processes as well.

How Do Computers Think?

That human beings can think is regarded as well established. If we wish to test this capability in any instance, we give the person a task of the kind that is regarded as requiring thought and observe his or her performance. That computers can think is sometimes regarded as debatable. If we wish to test this hypothesis, we give the computer a task of the kind used to test whether people think and observe its performance. The empirical evidence is by now overwhelming that computers can perform successfully in many task environments that call for thinking in humans. Hence we must conclude that appropriately programmed computers think.

The more interesting and difficult question is whether a thinking computer uses the same processes as a thinking person in the same task environment. The empirical answer is "sometimes yes, sometimes no." For example, there now exist some chessplaying computers that perform at a formidable level, so that they can be defeated by only a few hundred (at most) of the strongest human players. These programs demonstrably use processes that differ in important ways from those used by human players. While, like human players, they draw upon a considerable body of chess knowledge, they also conduct an enormous search through the tree of possible moves and countermoves (typically a search through several million branches at each move). Human players, including masters and grandmasters, seldom examine as many as one hundred branches in the game tree before making a move (de Groot, 1978).

On the other hand, several chess programs have been constructed that behave in a far more humanoid manner. A program built by Newell, Shaw, and Simon (1958) selected its moves by forming and pursuing goals, with a very small amount of search. It was a weak, but credible, chess player. A mating combinations

program by Baylor and Simon (1966) was very powerful in searching for winning moves in sharp tactical situations, usually examining far fewer than one hundred branches and finding many of the celebrated grandmaster combinations that are reproduced in histories of chess. These programs can be shown to exhibit many of the phenomena of human chess thinking -- tactical thinking more successfully than strategic thinking.

These investigations of chess playing can be matched by simulations of human expertise in many other task domains: solving puzzles, solving mathematics and physics problems, making medical diagnoses, detecting patterns in letter and number sequences, learning languages, understanding problem instructions, and others.

Principles of Human Problem Solving

Let me comment now on the content of the theory of problem solving that has emerged from this research. I will focus on four major principles, omitting many details.

First, most problem solving involves selective search through large spaces of possibilities. The selectivity, based on rules of thumb or *heuristics*, allows such searches frequently to reach success in a reasonable length of time, where an undirected, trial-and-error search would require an enormous time, and often could not be completed in a human lifetime. In cases of well-practiced skills, there may be almost no search at all, the heuristics being sufficiently powerful to select the correct path to the goal at once. The more difficult the problem and the less efficient the heuristics, the more search will be involved. (We will expect tasks demanding creativity to require a good deal of search, for by definition, the heuristics available will be weak.)

Second, some of the heuristics that guide search in problem solving are specific to the task domain, but others are quite general, applying to a wide range

of domains. Because of their generality, these latter methods cannot make use of information that is specific to a given domain, hence they are generally called "weak methods." When stronger, domain-specific heuristics are not available to a problem solver, we would expect him or her to fall back on weak methods. Therefore, we would expect weak methods to play an important role in explorations of new territory -- hence in scientific discovery.

Third, among the most important and widely used problem solving heuristics is means-ends analysis. The problem solver compares the present situation with the goal situation and notices a difference between them. Recognition of the difference cues in memory information about the operators that might be applied to remove it, bringing the situation closer to the goal. Once a new situation is attained, the same procedure is applied again, until the goal is reached. The effectiveness of means-ends analysis depends on the problem solver's ability to notice differences and to retrieve from memory relevant operators for reducing or removing these differences. Hence, in the presence of considerable domain-specific knowledge of this kind, means-ends analysis is a strong method; if there is little such knowledge, it serves as a weak method.

Fourth, the domain-specific knowledge that largely distinguishes expert from novice behavior is stored in memory in the form of *productions*, that is to say, of actions (A) paired to conditions (C), C ---> A. When the conditions are satisfied in a problem situation (e.g., when the problem solver notices the presence of certain cues), memory is accessed for actions that are associated with these conditions. (E.g., in the case of means-ends analysis, the conditions are what we called differences, and the actions are operators for removing these differences.) Each execution of a production is an act of recognition. Thus, the physician, recognizing a patient's symptoms, is reminded of one or more diseases that present these

symptoms and of the nature of these diseases, tests that will discriminate among them, prognosis and treatment, and so on.

The expert's skill, then, derives from a large base of productions (50,000 is a plausible number for human professional-level skills) that allow recognition of relevant cues in problem situations, combined with an ability to analyse and reason about information using means-ends analysis and similar heuristics. In this picture of expertise, there is no sharp boundary between "insight" or "intuition" and analysis. Insight and intuition are simply acts of recognition based on the stored knowledge of the domain, and they interact with analysis to solve problems. For many simple, everyday problems, recognition alone may be enough, and little analysis may be necessary. In more complex situations, recognition of salient cues allows analysis to take larger and more appropriate steps than if the heuristic search has to depend on weak methods alone.

The claim, then, that the processes of scientific discovery are normal problem-solving processes is a claim that scientific discovery follows the four principles just enunciated. First, its basic method is selective (heuristic) search. Second it uses both general and domain-specific heuristics. Third, means-ends analysis, a heuristic of broad applicability, plays an important role in analysis and reasoning. Fourth, effectiveness in discovery depends heavily on processes of recognition, making use of tens of thousands of productions that index memory with familiar and recognizable cues characterizing common problem situations.

The Processes of Scientific Discovery

Scientists, in their work, do a great many different things. In most sciences, there is a greater or lesser degree of specialization between theorists and experimentalists, but the specialization, formal and informal, goes much farther than

this. Scientists discover and define problems, they find appropriate representations for problems, they design experimental procedures and strategies and plan and execute experiments, they obtain data by observation, they formulate laws and theories to account for data, using mathematical and other forms of reasoning, they deduce consequences from their theories, they invent instruments for making observations, and they devise explanatory theories to give deeper accounts of descriptive laws.

This may not be an exhaustive list of scientists' activities, but it will serve to illustrate the wide variety of activities that make up the scientific enterprise. My claim is that all of these are problem solving activities that make use of the basic processes described in the previous section. Since we now have direct empirical evidence to support this claim for several, if not all, of these activities, I would like to provide some examples that will illustrate the nature of the evidence.

Deriving Laws from Data

The activity that has been studied perhaps most intensively is the process of deriving laws from data. A series of computer programs, collectively named BACON, has successfully simulated the process whereby a substantial number of important laws of eighteenth and nineteenth century chemistry were induced from data. (For a detailed account of BACON and related programs, see Langley, et al., 1987.) In all of these cases, we know the process was an inductive one, without benefit of theory, because there did not exist, at the time of the discoveries, theories that were relevant to the derivation.

For example, BACON is able to obtain Kepler's Third Law from data on the distances and periods of revolution of the planets about the Sun. The discovery process is purely inductive, as was Kepler's, since there existed in his time no body of theory that would have led to the law -- that theory had to wait for Newton,

generations later. The BACON program also obtains Joseph Black's laws of temperature equilibrium, the law of conservation of momentum, and others, and discovers atomic and molecular weights from data on chemical reactions, distinguishing between atoms and molecules along the way. It not only finds laws, but in a number of instances introduces new theoretical concepts (inertial mass, index of refraction, specific heat, atomic weight, molecular weight) to permit the laws to be stated parsimoniously. A detailed account of BACON's successes (and limitations) can be found in Langley et al. (1987).

Other programs, also reviewed in Langley et al. (1987), derive qualitative laws from qualitative data. The STAHL program, for example, on examining information about combustion and reduction reactions, will arrive at the phlogiston theory of combustion or the oxygen theory, depending on how the reactions are described to it. (The actual history of the competition of the phlogiston and oxygen theories can be explained in these same terms.)

Our knowledge of the induction of laws from data is not limited to computer simulations of historical cases. We can also present the historical data (unidentified) to subjects in the psychological laboratory, and examine their attempts to find regularities in the data. Qin and I (unpublished) have presented 14 students with Kepler's data; four of our subjects found Kepler's Third Law in one to two hours' work. When we compare the thinking-aloud protocols of the successful and unsuccessful subjects, we find that the former, but not the latter, compared each hypothesis they formed with the actual data and then selected their next hypothesis on the basis of the specific discrepancies they found. Their heuristics in generating a sequence of hypotheses closely resembled BACON's, as did the actual hypotheses they generated. Similar experimental data, with similar findings, have been gathered for Balmer's law, which describes the wave lengths of lines in the hydrogen

spectrum, and Planck's law of blackbody radiation, both foundation stones of subsequent quantum theory. Subjects were often able to find these laws in the laboratory, and the heuristics they used in the search were similar to BACON's.

Planning Experiments

Quite different from the activity of discovering laws in data is the process of planning sequences of experiments aimed at producing data relevant to a research goal. For example, the German biologist, Hans Krebs, carried out a series of experiments over about nine months in 1931-32 that revealed the reactions that synthesize urea in vivo. (For details of this history, see Holmes, 1980.) Of course, the experiments were not all planned in advance. On the contrary, each experiment provided information that led to the gradual (and in one case, sudden) modification of the research plan.

Krebs began with the decision to try to discover the process of urea synthesis in vivo, using slices of liver tissue as his experimental materials. The research problem was an important one, already well recognized in the field, that had not yielded to previously available methods of experimentation. The tissue slice method was a new one that Krebs had acquired while working as a postdoctoral student with Otto Warburg. Krebs' initial strategy was to repeat experiments that had been performed on whole organs (the method previously used) to see if he could reproduce their results. Many of these experiments involved testing the urea yield when a tissue was treated with mixtures of ammonia and an amino acid. The yields of urea were moderate until a particular amino acid, ornithine, was tested; the yield with ornithine was quite large.

Krebs now switched to a new strategy, which we might call the "response to surprise strategy." He first sought to determine whether other molecules similar to ornithine would produce the same high yield of urea. They did not. He then

proceeded to vary the quantities of ornithine and ammonia used and to measure the changing yields, while at the same time trying to work out, using his knowledge of the chemical structures of the reacting molecules, plausible reaction paths. He discovered such a path, in which ammonia provided the nitrogen for the urea, while ornithine served as a catalyst in the cyclical reaction.

Deepak Kulkarni has constructed a computer program, KEKADA, which captures many of the heuristics that guided Krebs' strategic planning. On the basis of its experience, it forms expectations about the outcomes of these experiments, and when these expectations are disappointed, it adopts a "respond to surprise" strategy that involves delimiting the scope of the surprising phenomenon and then searching for its mechanism (Kulkarni & Simon, 1988). Not too surprisingly, KEKADA does a good job of simulating Krebs' urea synthesis discovery. More impressive, when provided with the appropriate initial conditions (research problem, available methods, and domain knowledge), it also simulates closely (1) Krebs' discovery of the glutamine cycle, (2) some 19th century research on the synthesis of alcohols, and (3) Faraday's research strategy after his "surprise" in finding that changes in magnetic fields could induce electric currents.

KEKADA is able to simulate these disparate phenomena because most of its experiment-planning heuristics are independent of the precise task domain to which they are applied. The heuristics for responding to surprising phenomena are critical in accounting for its success.

The historical data on the discoveries of Krebs, Faraday, and others, together with the interpretation provided by KEKADA shows the experimental process of very successful investigators to consist in heuristic search through a large space of possible experiments, the heuristics guiding the selection of each successive experiment. Many of the basic heuristics are, like the heuristics for exploiting

surprise, quite general and independent of the specific task domain in which they are applied. Experimentation, it would appear, is a process of heuristic search resembling closely the processes that have been observed and identified in other kinds of problem solving.

Explanatory Theories

When we fit a mathematical function to data, as BACON does, we are at best providing a parsimonious description of the data. The function, the law, does not explain why the data are as they are. Thus, Kepler described the relation between the periods of the planets' orbits and their distances from the Sun; Newton explained the relation by showing that it followed logically from the inverse square law of gravitational attraction. Balmer found a formula to describe the successive lines of the hydrogen spectrum; thirty years later, Bohr showed that Balmer's formula could be deduced from his quantum model of the hydrogen atom. Science is as interested in discovering mechanisms that explain phenomena as it is in discovering laws that describe them. What discovery processes enable explanations to be found?

BACON, as applied to the phenomena of temperature equilibrium, throws some light on this question. Suppose we provide BACON with some very broad theoretical concepts: that when substances are mixed together both mass and heat are conserved; and that the law describing the temperature equilibrium of such a mixture should be symmetrical in the properties of the components. Then, Black's law of temperature equilibrium can be deduced from these assumptions, in advance of any examination of data, and the data simply used to confirm the law. Without these assumptions, BACON must induce the law from data with the help of rather arduous calculations. Thus assumptions of conservation and symmetry can be used as heuristics to reduce the search required to find laws. If the search is successful (with or without the heuristics), the heuristics then provide at least a partial

explanation of why the phenomena are as they are.

Explanation often requires us to consider the phenomena of interest at a more microscopic level than the level of observation. We postulate unobservable mechanisms, to account for the observable data. The DALTON program, for example (Langley et al., 1987), assumes that chemical substances are made up of molecules, and molecules of atoms of the elements. Further, it assumes that atoms are conserved in reactions and that volumes of gases (under constant pressure) are proportional to the numbers of their molecules (Gay-Lussac's Law). Starting with these assumptions and data about the inputs and outputs of chemical reactions, it deduces the chemical formulas of the molecules involved. For example, on being told that three volumes of hydrogen and one of nitrogen produce two of ammonia, it concludes, correctly, that hydrogen and nitrogen are H2 and N2, respectively, and that ammonia has the formula NH3.

These simple examples show how we can begin to understand the discovery of explanatary theories as a problem solving process. The process starts with a representation of the phenomena (in the DALTON case, a particulate representation; in the case of Black's law, a representation in terms of conserved quantities of matter and of a substance called "heat"). This representation imposes constraints upon the phenomena that allow the mechanisms to be inferred from the data -- or even inferred deductively in some cases.

The question remains open of where representations come from. Answers to that question are just now beginning to be sought, and I will have nothing say about them here. But you can guess what my prediction is about them: that representations are found by means of ordinary problem solving processes.

The Invention of Instruments

I will comment on one other facet of scientific discovery that has not yet been studied in as much detail as the discovery of laws in data, the planning of sequences of experiments, or the discovery of explanatory theories: discovery that consists in the invention of new scientific instruments.

New instruments are commonly the byproducts of the observation of new phenomena; and of course new phenomena are commonly the products of new instruments. How does this chicken-and-egg process proceed?

Consider the case of temperature and thermometers for measuring it. Sensations of heat and cold provide human beings with a built-in thermometer requiring no artificial instrumentation. These sensations do not provide, however, a quantitative and invariant measuring scale that could serve as foundation for the laws of heat. However, experiments on heating various kinds of materials revealed a common phenomena: that many substances, solids and gases, expand when heated. By using standard methods for measuring volumes to determine the amount of expansion, the thermometer was created, in many forms corresponding to different substances (Langley et al., pp. 313-314). This basic idea was successively refined -- for example, by using the thermometer bulb to magnify the effects -- to produce instruments that we still use today.

Soon after the thermometer was invented, we find Fahrenheit and Boerhaave, followed by Joseph Black, laying down the quantitative laws of temperature equilibrium. The phenomenon of expansion on heating permitted the invention of the thermometer; the thermometer permitted observation of new phenomena of temperature equilibrium. A similar story can be told of the invention of such instruments as the ammeter and voltmeter following on the discovery of electrical currents and their magnetic effects. These instruments, in turn, permitted Ohm to

find his quantitative law of the relation among current, voltage, and resistance.

Heuristic search can again account for these discovery processes. One heuristic suggests looking for instruments that make use of new phenomena; another, even more obvious, heuristic suggests using instruments to find new phenomena.

Intuition, Insight, and Inspiration

We cannot leave the topic of scientific discovery, however, without attending directly to some of the phenomena that are most commonly introduced into evidence as a basis for claims that discovery is, somehow, different from other kinds of problem solving. It is often argued that creative discovery depends on such processes as intuition, insight, and inspiration, and anecdotal evidence is frequently brought forward to show their essentiality. Poincare achieves an understanding of the Fuschsian functions as he steps onto the bus at Coutances, Kekule conceives of the benzine ring as, half asleep, he watches the twisting snake of the fireplace flames grasp its tail in its mouth, and so on.

The principal phenomena that support the claims for intuition, insight, and inspiration are the suddenness with which a discovery is sometimes made (often preceded by a long period of unsuccessful work followed by a longer or shorter interruption), and the fact that the discoverer often cannot explain why it occurred just then, or what path led to it. If the signatures of intuition and insight are suddenness of discovery and incomplete awareness of the discovery path, then these earmarks do not distinguish these two processes from the well-known and well-understood process that we call "recognition."

The ability to recognize particular symptoms, or stimuli, depends on their familiarity from previous experience and learning. Various models have been proposed for the recognition process -- for example, the EPAM model, which

assumes that long-term memory is indexed by a discrimination net, which sorts the presented stimulus to find, if it is familiar, the information associated with it in memory (Feigenbaum & Simon, 1984). An alternative model, the so-called PANDEMONIUM mechanism, does a similar sorting job, but achieves it by parallel rather than serial processing.

The process of recognition has long been studied by psychologists. An act of recognition generally takes about half a second, or longer. Of great importance, while a person is consciously aware of the result of the recognition process (is aware of what or who has been recognized), he or she is not aware of the process itself or the cues that were used to discriminate the stimulus. Recognition is "intuitive" in exactly the sense in which that word is used in the literature of discovery and creativity.

I discussed earlier the strong empirical evidence that an expert, in his or her domain of expertise, holds in memory some 50,000 different cues or symptoms that, when present in the situation, will evoke a recognition and consequent access to stored knowledge relevant to the cue. Each expert has 50,000 "friends" and extensive information about them. Compare this number with the 50,000 to 100,000 words that each of us has in the vocabulary of our native language. The evidence is compelling that the expert accomplishes most of his or her daily work by means of this capability for recognizing situations and thereby recalling the knowledge necessary for dealing with them. At all steps of problem solving, recognition is intermingled with analysis, and without it, analysis is hopelessly slow, faltering, and inefficient.

The recognition mechanism can account quite adequately for Poincare's sudden discovery as he boarded the bus. It does not, by itself, explain the possible role of interruption or incubation, but simple explanations have been provided for these also

(Simon, 1977, pp. 292-299). The visual aspects of Kekule's experience (which, by the way, was first reported by him thirty years after the event) call for other mechanisms, but have nothing directly to do with the suddenness of the discovery or its subconscious origins.

Intuitions, insights, and inspirations are not only sudden, but they are also frequently surprising. In our analysis of KEKADA, we have seen that surprise is simply a form of recognition -- recognition that one's expectations have been disappointed. To have expectations, one must have knowledge as to what to expect. As Pasteur put it, "Accidents happen to the prepared mind." So again, we come back to the expert's 50,000 chunks that allow a recognition that something unusual has happened.

In summary, we do not need to postulate special mechanisms to account for intuition, insight, or inspiration. These phenomena will be produced by the mechanism of recognition, which we have already seen plays a key role in every form of expertise, and which is based, in turn, on the store of indexed knowledge that every expert possesses.

Conclusion: the Processes of Discovery

This quick and highly incomplete account of the evidence now available about discovery processes confirms both the variety and heterogeneity of the activities that make up the enterprise of science and the consistency with which these activities conform to the pattern of heuristic search -- highly selective search that produces some measure of success even in large and poorly structured problem spaces. Such phenomena as intuition, insight, and inspiration derive from the capacity for recognition that every expert acquires in his or her domain of expertise. No new mechanisms need be postulated to account for them.

Of course there is much room for additional study of the processes of science, and some processes, like problem representation, have hardly been touched by research to date. New research may certainly produce surprises, which will no doubt evoke the "respond to surprise" heuristic, leading to a different picture of the process. Each person can estimate his own prior probability, based on the evidence to date, that the theory of discovery will or will not be altered in fundamental respects.

References

Baylor, G. W., and H. A. Simon, A chess mating combinations program (1966), reprinted in Simon (1979), Chapter 4.3.

De Groot, A., *Thought and Choice in Chess*, The Hague: Mouton, 2nd Edition, 1978.

Ericsson, K. A., and H. A. Simon, *Protocol Analysis*, Cambridge, MA: MIT Press, 1984.

Faraday, M., *Faraday's Diary*, 8 vol., London: Bell, 1932-36.

Feigenbaum, E. A., B. Buchanan, and J. Lederberg, On generality and problem solving: A case study using the DENDRAL program, in B. Meltzer and D. Michie (eds.), *Machine Intelligence* 6, New York: Elsevier, 1971.

Feigenbaum, E. A., and H. A. Simon, EPAM-like models of recognition and learning, *Cognitive Science*, 8:305-336 (1984).

Giere, R. N., *Explaining Science*. Chicago: U. of Chicago Press, 1988.

Glass, A. L., K. J. Holyoak, and J. L. Santa, *Cognition*, Reading, MA: Addison-Wesley, 1979.

Gruber, H. E., *Darwin on Man*, New York: Dutton, 1974.

Kulkarni, D., and H. A. Simon, The processes of scientific discovery: the strategy of experimentation, *Cognitive Science* 12:139-175 (1988).

Langley, P., H. A. Simon, G. L. Bradshaw, and J. M. Zytkov, *Scientific Discovery*, Cambridge, MA: MIT Press, 1987.

Holmes, F. L., Hans Kerbs and the discovery of the ornithine cycle, *Federation Proceedings* 39:216-225 (1980).

Newell, A., J. C. Shaw, and H. A. Simon, Chess-playing programs and the problem of complexity, *IBM Journal of Research and Development*, 2:320-335 (1958).

Newell, A., and H. A. Simon, *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall, 1972.

Nickles, T., *Scientific Discovery*, Dordrecht: Reidel, 1978.

Popper, K. R., *The Logic of Scientific Discovery*, New York: Science Editions, 1961.

Qin, Y. L, and H. A. Simon, Laboratory replication of scientific discovery. Technical Report, Department of Psychology, Carnegie-Mellon University, 1988.

Simon, H. A., *Models of Discovery*, Dordrecht: Reidel, 1977.

Simon, H. A., *Models of Thought*, New Haven: Yale U. Press, Vol. 1, 1979; Vol. 2, 1989.